Hard choices and weak wills: The theory of intrapersonal dilemmas

Read, D; Roelofsma, P.H.M.P.

published in
Philosophical Psychology
1999

DOI (link to publisher)
10.1080/095150899105819

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Read, D., & Roelofsma, P. H. M. P. (1999). Hard choices and weak wills: The theory of intrapersonal dilemmas. *Philosophical Psychology*, *12*(3), 341-356. https://doi.org/10.1080/095150899105819

PHILOSOPHICAL PSYCHOLOGY, VOL. 12, NO. 3, 1999

Hard choices and weak wills: the theory of intrapersonal dilemmas

DANIEL READ & PETER ROELOFSMA

ABSTRACT *Social dilemmas occur when individuals make choices that are in their own best interest but not in the interest of society as a whole. Intrapersonal dilemmas occur when people make choices that are in the best interest of themselves at the moment of choice, but not in the best interest of themselves in the long run. A number of writers have observed that we can usefully model this self-defeating behavior by treating each individual as an aggregate of "selves" which have competing interests, like the individual selves in a population. We undertake to synthesize and extend these contributions, applying the theory of intrapersonal dilemmas to addiction and procrastination, and discuss how the concept of multiple selves can explain why people attempt to control their behavior by imposing costs or constraints on themselves in the future, and how it can also help us develop new strategies of (personal) behavior control.*

We begin by introducing you to a friend whose tragic life presents a lesson for us all. While young, he fell prey to temptation and adopted deleterious habits. Now he smokes heavily, drinks to excess, eats fattening foods, and engages in further depravities which decorum prevents us from disclosing. Every aspect of his life has suffered from his excesses. His marriage is on the rocks, he is constantly in trouble with the law, and he publishes very little. He knows that he should relinquish his ruinous lifestyle, but whenever he chooses between further indulgence and a moderate alternative, indulgence invariably wins out. Yet there is little pleasure in it all—cigarettes have long since gone from being stimulants to mere correctives for deprivation, alcohol is an antidote to memory, and, every day, the sixth cup of coffee is far less enjoyable than the first. As a result of his indulgences, our friend has fallen into decrepitude, yet those very vices which led to his deplorable state remain more pleasurable than the virtues which might have prevented his decline, and could yet reverse it.

Our poor friend's personal tragedy is similar to another one, described vividly and famously by Garrett Hardin in 1968. Hardin spoke of the tragedy of the *commons*. He introduced us to a bucolic village in rural England in the days before the laws of enclosure. Each village had a common area which could be used by each

Daniel Read (corresponding author), Leeds University Business School, Leeds LS2 9JT. Email: dr@lubs.leeds.ac.uk; Peter Roelofsma, Free University, Amsterdam and Leeds University Business School.

ISSN 0951-5089 (print) ISSN 1465-394X (online)/99/030341-16 © 1999 Taylor & Francis Ltd

villager according to his or her needs. Everybody recognized that they could better themselves personally by grazing cattle on the common land, thereby augmenting their herd without having to acquire new pasture of their own. These commons quickly became overburdened with cattle, and therefore became useless for anything other than producing poorly nourished cattle. Although as a whole the community got less out of each additional cow than did the villager who placed it there, that villager got more from the cow than he or she lost from the damage to the common land.

The tragedy of the commons is a tragedy of many people acting in their individual best interest and thereby undermining the interests of all. The individual players in this tragedy could all choose between an alternative which was socially desirable but individually undesirable (a *social* good), and another alternative which was socially undesirable but individually desirable (an *individual* good). In the end all chose the individual good, even though they would have been better off if all had chosen the social good [1]. Our friend's tragedy arises from an intrapersonal equivalent of the commons dilemma. He faces many choices between alternatives which are good for him in the long term, but not so good in the short term (these are *virtues*), and alternatives which are pleasant in the short term but have to be paid for later (*vices*) [2]. By choosing a vice our friend does the intrapersonal equivalent of grazing another cow on the commons. For the rest of his life he will experience greater hardship than if he had chosen virtue. If he could act on the basis of the "big picture" and choose for his life as a whole, he would always choose virtues over vices, just as one of Hardin's villager would always choose a pristine commons over a plundered one.

The idea that our own problems of self-control are like social dilemmas is not new (Ainslie, 1992; Schelling, 1992; Herrnstein & Prelec, 1992). The present paper synthesizes these contributions as well as providing our own perspective. We begin by laying out the two cornerstones of a theory of intrapersonal dilemmas. The first cornerstone is that the individual actor is a bundle of egoistic "selves," all of which put more emphasis on their own desires than on the desires of their compatriots (we call these *personal*, as opposed to *social*, selves). The second cornerstone is that the utility experienced by each personal self can be added to the utility experienced by other personal selves, and that the total utility (integrated over all personal selves) from different courses of action can compared. In this way, it is possible to establish that some decisions are objectively better than others for the entire population of selves, even when those decisions may not be the best for some personal selves, including the one making the decision.

Multiple selves

Social dilemmas occur when the interests of those who make the decisions differ from those who experience their consequences. For intrapersonal decisions to be like social dilemmas, there must be a way to subdivide individuals so that, theoretically speaking, they are "different people" within the same skin. It is also necessary that these different people—or selves—have interests that sometimes conflict. Such

multiple-self models have a long and distinguished history in the human sciences (Elster, 1986). Sometimes the selves coexist, and decision making results from a struggle between different interests. One such partitioning is Freud's tripartite division of the self into an id, ego and superego. Another, holds that these battling selves are the different social roles (e.g. teacher, friend, father) which the person can play. Sometimes, one self is a dictator and the others can only look on. For instance, Thaler and Shefrin (1981) distinguish between a *doer* who has the ability to act, and a *planner* who works out what the doer "should" do. Of course, the doer often disregards the desires of the planner.

One of the most widely applied multiple-self models treats the individual as a sequence of selves distributed over time, with each self taking the baton, as it were, from its predecessor. The wishes and desires of the currently active self receives special status when decisions are being made. Some consideration may be given to the desires of future selves, but the natural disposition is to give them little weight. Ainslie (1975, 1992), whose work has been most influential, refers to these selves as *successive motivational states*, and argues that they arise because of *hyperbolic discounting* [3].

Hyperbolic discounting was first proposed by learning theorists like Mazur (1987) and Herrnstein (1997) and has had great influence outside of its original domain, especially in philosophy (e.g. Nozick, 1993; Bratman, 1996; Mele, 1996; Wilson, 1995), and economics (Laibson, 1997; O'Donaghue & Rabin, in press). The concept is simple, but if it actually describes our preferences (and, to a first approximation, it probably does) then it has profound consequences.

Hyperbolic discounting for two rewards that differ in magnitude and delay is depicted in Figure 1. The vertical lines topped with circles represent two rewards. The *Y*-axis depicts utility, and the *X*-axis depicts time, or the delay to the receipt of each reward. The two curves which intersect with the top of the two reward lines show the "present value," in utility terms, to the decision maker of each reward for a given delay. One of the rewards is larger but comes later, while the other reward is smaller but comes sooner. The larger-later reward is a stylized *virtue* (e.g. heaven or comfortable retirement), while the smaller-sooner reward is a *vice* (carnal pleasure or a yuppie lifestyle). Notice that the two value curves intersect at one point, labeled the *indifference point*. Before this, the virtue is preferred, but once the threshold has been passed, the vice is preferred. These preference reversals occur because of an immediacy effect—the immediate future is given disproportionate weight in decision making.

Notice that the crossover point effectively divides the decision maker into three selves, which we have labeled A, B and C in Figure 1. Self A, before the crossover point, may have relatively little at stake in the decision—because the two alternatives are so remote, neither of them are valued very highly. If she does choose, however, she will take the virtue because it has a higher present value. Once the crossover point is passed, however, we can say that the vice is "present" and the overweighting of that present now leads the decision maker to prefer it. Finally, once the vice is no longer available, there is a third period in which the decision maker regrets her

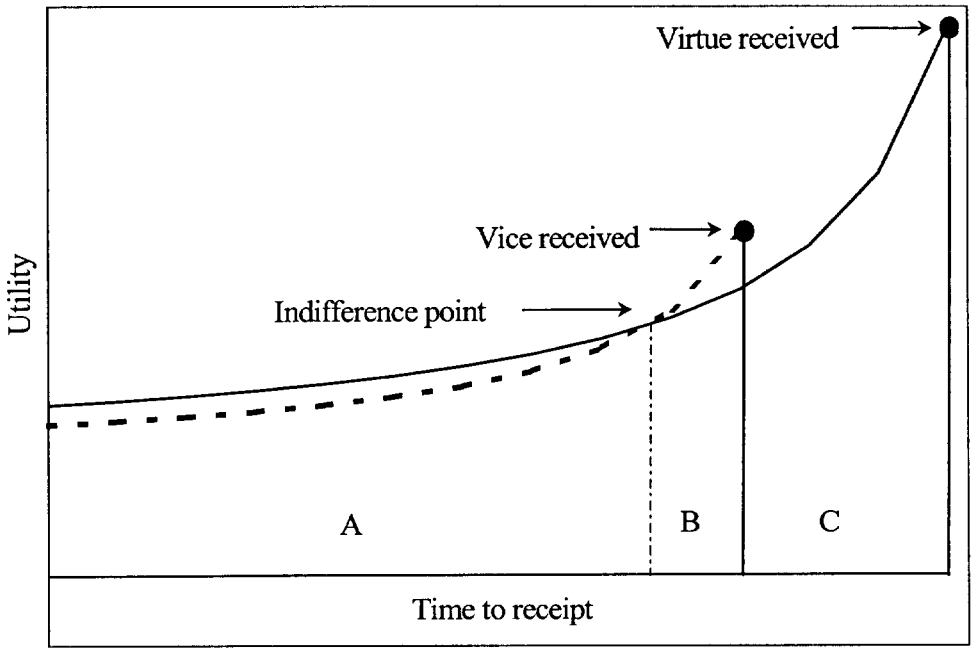


FIG. 1

impulsive choice (if she chose the vice) or feels relief over her "self-control." The doer and planner can also be located in this analysis. The planner is the self before the crossover point, who wants to take the virtue. The doer is the one who actually makes the decision, and that is usually the self who wants the vice. As we discuss below, planners sometimes act to subvert the expected shortsightedness of the doer by trying to put temptation out of the way or changing the payoffs so that the vice is rendered less attractive (Schelling, 1984a,b).

We have introduced some technical terms concerning time discounting to describe how individuals can be said to consist of multiple selves. The similarities between social and personal selves do not end there, however, and we briefly mention some qualitative similarities. We have already mentioned that social selves give special consideration to their own needs over those of others, just as individuals give special consideration to their current needs over their future ones. Social selves are rarely completely selfish and are usually somewhat concerned about one another, in much the same way as present personal selves are not completely indifferent about future personal selves. Social selves obtain utility from good outcomes that occur even to remote strangers, and the level of that concern increases with the degree of relatedness. We are closest to our kin, and get more utility out of their happiness than we do from the happiness of strangers; J.B.S. Haldane famously said that he would lay down his life for two brothers or four cousins (cited in Smith, 1975/1993). These close relatives are just like our immediate temporal descendents [4]. The closer in time they are to ourselves, the more we care about them.

Personal selves are also like social selves in that they have wishes concerning

what past selves might have done, and hopes about what future selves will do. Imagine present selves who are choosing between vice and virtue. As like as not, they will choose vice. They will also know whether they chose vices or virtues in the past. Because present selves would be better off if their past selves had chosen virtue, they will wish that they had chosen virtue, although realistically it is likely that they chose vice. Likewise, because they have some concern for their future selves, and because they know that virtues are better for them than vices, they will hope that these future selves will choose virtue, although they are likely to expect (or fear) that they will choose vice. This characterizes the smoker who wishes that he didn't smoke (i.e. that past selves hadn't chosen cigarettes), hopes that he won't smoke in the future, but lights up right now. Most of us have seen, if not actually been, the smoker who resolves firmly but vainly that *this* is his "last cigarette."

So far, we have spoken of the degree to which the current self values, or takes into account, the needs and desires of future selves. We can refer to this as *intrapersonal sympathy*. In order for social choices to be optimal, however, it is not only necessary for people to care about what others want, but they must also know what they want. Indeed, it is commonplace that social problems are often caused by a failure of groups or individuals to understand the feelings of others. People talking in the cinema, for instance, may well care about the pleasure of their neighbors, and yet not realize that they are destroying it. Likewise, if our current selves do not know what our future selves will want or do, we may end up inadvertently "offending" them, or putting them in compromising situations. Loewenstein (1996) recently introduced the term *intrapersonal empathy* to refer to our ability to understand how past selves felt, and how future selves will feel. He argued that many situations—especially those in which strong emotions or desires are involved—are characterized by intrapersonal empathy *gaps*, in which a present self (we use our terms here) cannot put itself in the shoes of the future self for which it is deciding (Loewenstein *et al.*, 1997; Loewenstein, in press). A related issue was raised by O'Donoghue and Rabin (1997, in press), who distinguished between *sophisticated* decision makers, who know that their desires will change in the future and can thus act to either thwart them or facilitate them, and *naïve* decision makers who believe either that their desires will be constant or that they will have sufficient willpower to avoid succumbing to them. If a smoker vows to quit but probably won't, and if he knows that he probably won't quit, then he is sophisticated. If he thinks he will quit, he is naïve [5].

In order to have a social dilemma, there must be a society of people, and there must be a failure of the marketplace (Schelling, 1978). These market failures are called externalities. In this section we discussed the intrapersonal equivalent of society, and in the next section we discuss what we need in order to have a failure of an internal marketplace. First, we need a currency in which the internal marketplace operates, and then we need something akin to an externality.

Additive utility and internalities

In order to say that some people are "better off" choosing Y over X, even though they want to choose X and actually do choose it, we need to distinguish between

what people want now, and what they *should* want. For most readers, such a distinction will be uncontroversial. Most will agree that, for instance, even if the medicine tastes bad we should take it. More importantly, most of us would probably agree that if the medicine tastes so bad we will not take it voluntarily, we should find a way to trick ourselves into taking it, or somebody who will force us to take it, even if they have to find us and we hope they don't. One way of restating this is that the utility of taking the medicine is greater than the utility of not taking it.

Such a viewpoint is inconsistent with modern economic theory, which holds that the utility of an action is revealed by one's preferences. In other words, if you have a choice between medicine and no medicine, and you choose no medicine, then not taking the medicine has more utility than taking it. Clearly, based on this definition of utility we could never say that people sometimes choose X when Y would have given them more utility.

Utility originally had a different meaning, one that Varian (1996) decries as "old-fashioned" and unworkable. It referred to the pleasure or pain arising from an experience, and consequently would not necessarily be reflected in one's choices. Kahneman *et al.* (1997) have undertaken to revive this Victorian notion of utility in a paper appropriately entitled "Back to Bentham?" They argue that we can distinguish between *decision utility*, which is utility as conceived by modern economists such as Varian, and *experienced utility*, which is the utility that we (as consumers) usually care about—the happiness that we get from what we choose. Kahneman *et al.* (1997) go further, and argue that we can add up the *total* utility that comes from our choices. This is done by measuring the pain or pleasure experienced by a person at each moment, and then summing up all the measurements. An alternative is better than another if it has a higher total utility.

We can use the idea of instant utility to more precisely define what virtues and vices are, and to gain a different perspective on preference reversal. In Figure 2(a) the instant utility of two alternatives at a point in time is represented by the height of a line at that point. The total utility of each alternative, represented in Figure 2(b), is the area under the line integrated over the entire period. We assume, for the sake of simplicity, that the two alternatives yield utility only over the time period represented under the line, and we also assume no time discounting. The vice yields more utility than the virtue in the early periods, but less utility in the later periods. The total utility of the virtue is greater than that of the vice. If the agent gives more weight to early periods than later periods, they may end up choosing the vice. An important point is given by the vertical dotted line in Figure 2(a). If we integrate the area under the instant utility line for both virtues and vices up to this point, the total utilities are equal. Below this point, therefore, the vice provides more utility than the virtue, and above this point the virtue gives more utility than the vice. This gives us another division between selves, formerly discussed in terms of hyperbolic discounting. The current self will prefer the vice so long as their vision embraces a period shorter than that from the present to the dotted line.

If an agent chooses a vice instead of a virtue, they are imposing a cost on their later selves equal to the difference between the total utilities of the two alternatives. This cost has an analogue in the economics of social selves in the concept of

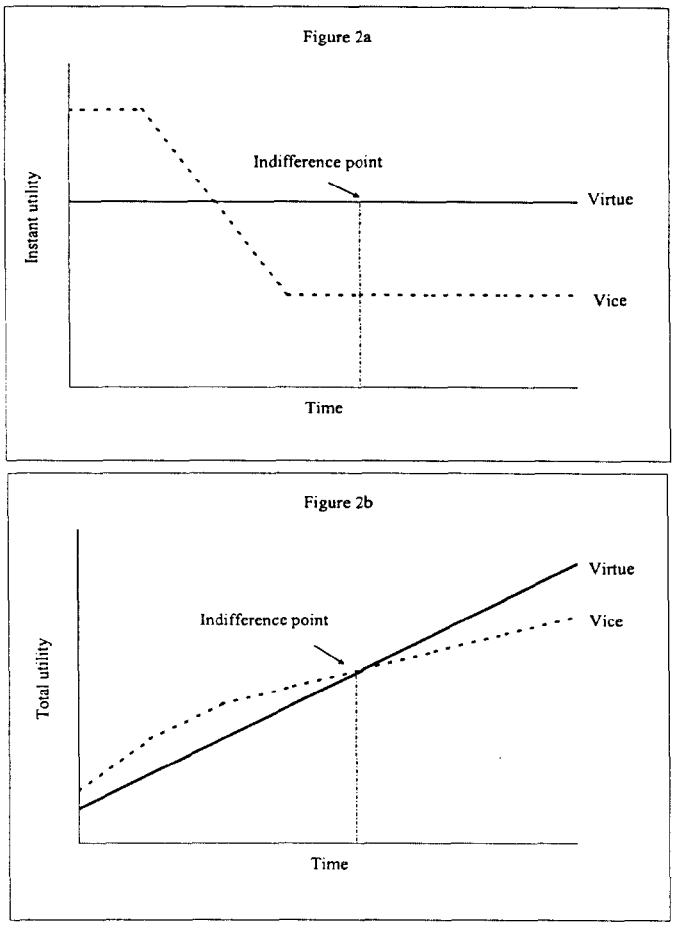


FIG. 2

externality (e.g. Schelling, 1978). Externalities are costs or benefits that are suffered or enjoyed, not by the individual decision maker, but by other members of the society. A classic example of a negative externality (a social cost) is the pollution caused by a car's exhaust. The driver of each car does not pay directly for his or her pollution (for instance, it is not pumped directly into his car) and so the individual decision to drive or not is unlikely to be much influenced by the pollution emitted from the car. This is a negative externality, but positive externalities are equally important. For instance, putting a bus on the road increases the amount of the road available to cars in proportion to the number of people who ride the bus instead of driving. The car drivers do not pay for this convenience. If bus riders were paid for the positive externality they create, and if car drivers paid for their negative externality, there would be fewer cars and more buses on the road. If the rate of payment and compensation were carefully chosen, social welfare would be maximized when everybody acted according to their self-interest.

If the agent is composed of a sequence of self-interested selves, then decisions made by the current self can bequeath costs or benefits to future selves. For instance, the decision to smoke now is a decision to impose a preference on future selves (due to nicotine dependency) that they may wish they did not have, as well as a poorer than possible health state. If the current self does not have to experience these effects, then they can have the same function for the stream of selves as externalities have for society. Herrnstein *et al.* (1993) coined the term *internality* to characterize internal costs that are ignored by the decision maker. Intrapersonal dilemmas, therefore, are caused by the same factor as are social dilemmas—the absence of an efficient internal market. If there was some way of taking these distributed costs and charging our present selves "in full," they would no longer be dilemmas. Internalities do not only have to be positive. A self that chooses not to do something that it would enjoy now in order to bequeath greater pleasure to its future self produces a positive internality which the future self cannot compensate it for.

Intrapersonal dilemmas take two common forms. First, the dilemma can arise when current choices affect future tastes, so that what is now enjoyable becomes either less or more enjoyable in the future because of current decisions. If these effects are ignored when making a choice, the internality is the difference between the future desirability of an option and the desirability it would have if the current choice was different. We call these *taste change dilemmas*. Second, a dilemma can arise because some of the costs or benefits that we gain from choosing an alternative are too small to be worth taking into account when they are considered individually. Again, when these small consequences are ignored they become internalities. We call these *adding-up dilemmas*. We next consider two "case histories," which exemplify each form.

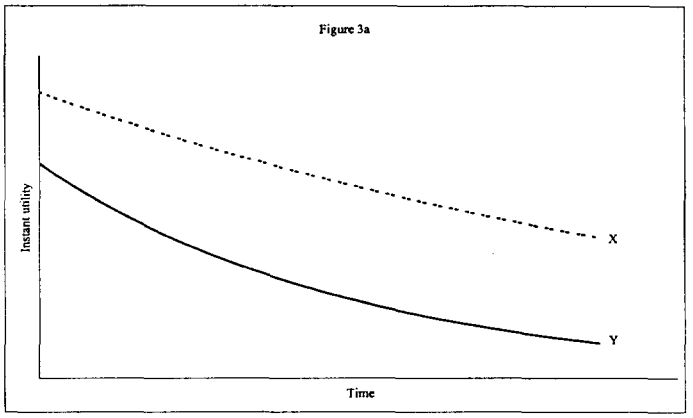
Addiction, the paradigmatic taste change dilemma

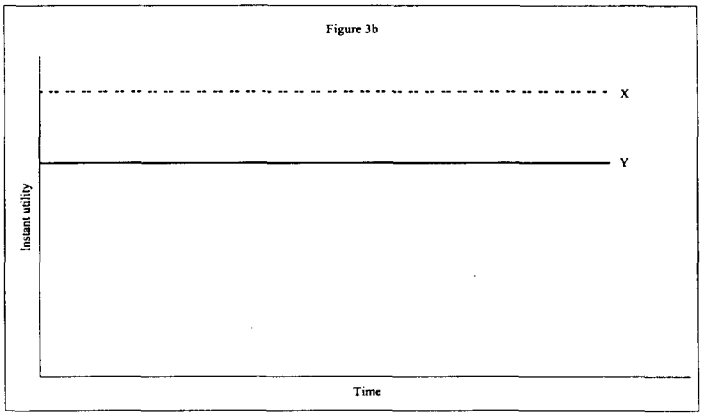
Nobody sets out to become addicted, but they become addicted nonetheless. Through using an addictive substance [6], an addict's tastes change in two ways. First, the quantity of addictive substance or experience needed to achieve or maintain a given level of happiness increases with repeated consumption. Gamblers, for example, need increasingly large and risky propositions to maintain the "thrill," as do habitual users of cocaine or caffeine. Second, the pleasure to be gained from other parts of the addict's life decreases over time, so that even though the addictive good is losing its power, it always remain more attractive than its alternatives. The gambler spends all of his time at the track or casino, and loses all interest in home or work [7]. Using the concept of total utility, Figure 2 depicts a stylized addiction dilemma [8]. At each point, the decision maker has two courses of action: choose the addictive (X) or non-addictive (Y) good. At time 0, X is preferred to Y [in Figures 3(a,b) the Y-axis shows the instant utility gained from each choice]. Figure 3(a) shows what happens if X is chosen every time. With each choice of X, the liking for both X and Y decreases. In our illustration, the proportional rate of decrease is constant for both X and Y, but the rate at which Y declines is greater than the rate at which X declines. This is characteristic of how addicts' tastes change: they

typically become obsessed with their addiction, and lose their taste for almost anything else. At the same time, the amount of the addictive substance or experience needed to obtain the same level of pleasure also increases. A gambling addict, for instance, may reach a point where constant gambling is necessary merely to maintain a minimum level of well being. Figure 2(b) shows what happens if *Y* is chosen every time—the value of both *Y* and *X* remain unchanged with *X* being eternally tempting. Figure 3(c) shows the total utility of the two courses of action depicted in Figures 3(a) and (b). Choosing *X* all the time yields a higher total utility for a while, but in the end the person is better off choosing *Y*. Figures 3(a) and (b) represent the two extreme courses of action, and intermediate cases are possible. For instance, imagine that the decision maker was able to choose *Y* until they reached a point where the amount of life remaining to them was less than the period from 0 to the dotted line, and then turns to *X*. In such a case, the person's total lifetime utility would be greater than that which he or she would get from a lifetime of choosing *Y*. This is because the internality from choosing *X* is borne by later selves. When there are no later selves (i.e. after death), there is no cost.

Addiction has a further element suggested by this analysis. Once addicted, it is very hard to become unaddicted (Kirby, 1996). This is why Figure 3(a) depicts the rate at which *Y* declines as being greater than the rate at which *X* declines. If the addict attempts to reform, he or she has to jump from the *X* line to the *Y* line at whatever point he or she finds himself. Once that jump is made, repeated choice of *Y* may lead one back to the path of normality, but it is not as easy to return from a point of addiction as it was to get there in the first place. The climb back along the *Y* alternative may be much slower than the fall, and for a long time (perhaps forever) *X* is a strong temptation. Lapsing by returning to *X* means the climb must start again, this one further hindered by the knowledge of past failure.

Taste change dilemmas do not always involve negative effects on future selves. When the taste changes at issue will *improve* the person in the future, a dilemma occurs if the current self would prefer not to make an investment (for which he will not be compensated) in the pleasure of the future self. A classic case is learning to play a musical instrument. For a novice, almost anything is more pleasurable than playing the violin, and so a large investment of time is needed to reach the point where it is an end in itself. We have observed that every one of our colleagues who plays a musical instrument was trained as a child (one of the authors is in this category), and everyone who has attempted to learn music as an adult has failed (the other author is in this category) [9]. Adults don't have parents standing over them while they practice, and learning to play an instrument is *very* hard.





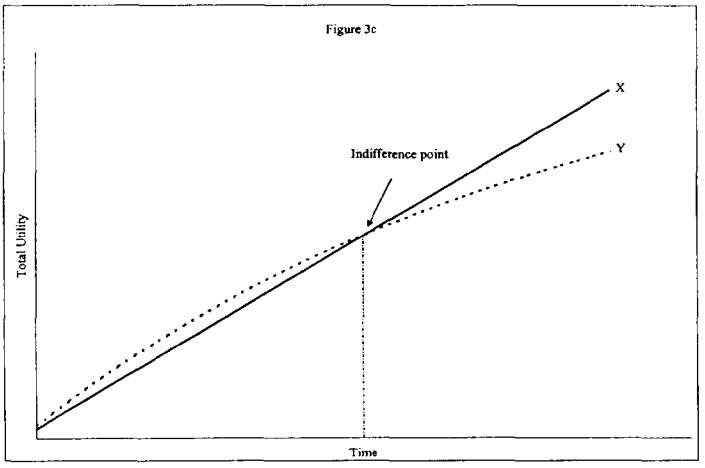


FIG. 3

Procrastination

In literature, as in the arch-procrastinators Hamlet and Oblomov, procrastination is typically portrayed as the result of indecision. In reality, however, procrastination is more likely to be the result of deferring the pain of acting onto a future self. Consider the following common scenario. Mary has borrowed a book from the local lending library, and it is now due. Fines are one dollar per day. Because the long trip to the library is not worth making just to save a dollar, Mary postpones returning the book until tomorrow. The next day she already owes one dollar, and if she doesn't return the book she will owe two dollars. But she can only save *one* dollar by returning the book (the original dollar is lost forever) and since it still isn't worth a dollar to make the trip to the library, Mary again fails to return the book. After 10 days or so she owes 10 dollars, but the *marginal* cost of not returning the book (that is, the cost for delaying it one day) is unchanged at one dollar. If she had realized that it was going to cost 10 dollars to return the book, Mary would have done so on the first day, yet even now she has no compelling (economic) reason to return the book today. The book will now remain unreturned, perhaps until the library grants an amnesty.

Some version of this story is frequently repeated in more poignant form. How many of us have lost touch with an old friend because we put off contacting him or her until so much time had passed that it seemed "too late"? Akerlof (1991) vividly recounts how he procrastinated for months over sending a package from India (where he was living) to a friend in the US until he finally gave up altogether and waited until he returned home when he could return the package in person. Mary procrastinates because each self in the series does not want to shoulder the burden of returning the book in order to forego a miniscule cost. However, these costs add up, and later selves have to shoulder the burden of all previous failures to act. There are many actions and failures to act that similarly ignore small costs and benefits that accumulate. Some familiar ones are smoking (our health declines with each cigarette, although we don't notice the change), not brushing our teeth, and eating dessert every day.

There is a close relationship between procrastination and a familiar social dilemma. Consider the following incident recounted by Thomas Schelling:

> Returning from Cape Cod on a Sunday afternoon, motorists were held up for a mile or more, at a creeping pace, by a mattress that had fallen off the top of some returning vacationer's station wagon. Nobody knows how many hundreds of cars slowed down a mile in advance, arrived at the mattress five minutes later, waited for the oncoming traffic, and swerved around before resuming speed. (Schelling, 1978, pp. 125–126)

Here is an undesirable social event. As Schelling observes, it would cost one of the drivers a few minutes of his or her time to remove the mattress from the road, and the driver would certainly have rather removed the mattress than wait in the traffic jam, but once they have reached the mattress their waiting time is over. Then it only costs a few seconds to drive past the mattress, much less time than it would take to remove it. Mary's future selves are just like the drivers in the queue. Each self who

does not return the book adds to the burden of later selves, just as each driver in the queue preceding any given car adds to the waiting time for that car. The lengthy wait of each driver is like the accumulation of fines. The more selves that have preceded the current Mary to the day when she has to decide whether or not to return the book, the greater the fine that she must pay. If just one of those selves (preferably an early one) had returned the library book, the future selves would not have had to bear the quite imposing fine. Just as the drivers will curse the indolence of those who went before them while they swerve around the mattress, so will Mary's current self fervently wish that one of those earlier selves had returned the book even though she is still not willing to do it herself.

We have already raised the issue of intrapersonal *empathy*, referring to our beliefs concerning what future selves will do or want. The problem of procrastination may arise because we believe that tomorrow's self will have less to do or will otherwise be more enthusiastic about going to the library than we are today (e.g. O'Donoghue & Rabin, 1997). This problem of thinking that tomorrow's self will be more virtuous than today's has not been studied very much (although see Read *et al.*, in press b, for a discussion and an example), but we suspect that it is a fruitful area for further investigation.

Overcoming intrapersonal dilemmas

Sometimes it is enough to recognize the dilemma in order to overcome it. Take the case of neglecting to return the library book. Usually, once we understand that if we don't return the book today we won't want to tomorrow and so forth, we recognize that we should return it today, and usually we will. However, this is not generally true even for procrastination, and it is certainly not true for many of the traps in which we find ourselves. We commonly find ourselves procrastinating over important tasks like writing papers. We know we are procrastinating, and that if we don't get down to work we will just suffer more tomorrow, but we procrastinate nonetheless. We procrastinate because work is difficult, and almost any other activity is more appealing. Sometimes, in fact, it seems impossible to not procrastinate, and we might say that we have "writer's block." In these cases we seem to be "out of control," in just the same way that a drug addict is out of control (Loewenstein, 1996).

The fact that we are sometimes overwhelmed by vice is what makes self-control so difficult. Heyman (1996) has argued that addiction (and by extension, any intrapersonal dilemma) can be overcome by putting ourselves under the control of *overall value functions* as opposed to *local value functions*. Overall value functions are our long-term interests. Heyman is undoubtedly correct that once this is done the problem will be solved, but the greatest part of the problem is finding out how to put ourselves under the control of these wise value functions. Although some alcoholics, drug addicts and procrastinators don't know that they have a problem, a fair number do. They are aware of what would be best for them in the long run, and they dearly want to do it. Most of them are probably also sophisticated—they know that what they are doing now is probably what they will do in the future. They know what their

overall value function is, and they want to bring themselves under its control, but they can't.

Self-control can come from the inside, or it can come from the outside. Usually, it comes from both. St Paul, an early multiple-self theorist, told the Corinthians that "I beat my body and make it my slave" (1 Corinthians 9:27, New International Version). This is the internal solution, in which one self tells the other self what to do. When this doesn't work, the solutions that people find to control their own behavior are remarkably like the solutions they find to control the behavior of others.

One solution is precommitment. Academics all over the world have marveled at one of the greatest inventions of all time—the deadline. In the absence of a deadline, many papers would remain unwritten. With a deadline, especially one that is fixed and immutable, a paper magically gets produced. Deadlines work by forcing us to prioritize (writing *now* becomes more important than fishing), and by imposing costs on not getting something done. It is almost unimaginable to fail to meet a deadline. Work that can take weeks will take days if there are only days in which to do it. Deadlines are sometimes imposed from outside (the tax form must be franked on April 15), but sometimes we impose them on ourselves. However, this can only work if there are some consequences for not meeting the deadline, and it is here that the academic conference works such wonders. There are dozens of conferences every year, and so it is always possible to find one that fits our special requirements.

Strotz (1956) offered an example of precommitment that, when we first read it, seemed contrary to the spirit of self-control. "The worker whose income is garnished chronically or who is continually harassed by creditors and who, when one oppressive debt is paid, immediately incurs another is commonly precommitting." To cite an example, imagine someone who constantly spends beyond their means, buying televisions and stereo equipment that stretches their budget to the breaking point. This person may well be acting impulsively, but by so doing they spend their money on durable goods, something that they might not do otherwise. Unfortunately, putting oneself into a legally enforceable debt is one of the few ways to force oneself to save.

A potentially more drastic solution, since it is premised on the fact that self-control is impossible, even with a deadline, is physical rationing or denial. Wertenbroch (in press) has shown that people buy vices in small quantities because they fear that if they buy more they will consume more. The most drastic version of this involves complete denial, or tying our hands. One of us has the good fortune to live in an apartment in which the TV cable is detachable from the wall. He keeps the cable in the office, and occasionally, for a treat, brings it home. Because cable TV is the only reception available, he literally cannot watch TV without making a trip to the office. People often refuse to buy cookies or magazines because they know that they are incapable of rationing them. Restrictions on judges (such as minimum sentencing laws) and legislators (balanced budget amendments) are designed to prevent these actors from giving in to their natural impulse to be lenient or to dig too deeply into the trough.

A drastic act of self-control sometimes occurs immediately after a vice has been consumed. The virtuous self gets the upper hand when the more carnal self has been

satisfied and the strength of desire is momentarily diminished. This is what William Faulkner, the poet of self-control and its failures, refers to as "rather that virtue which is desire's temporary assuagement than its permanent annealment." In the passage from *Pylon* from which this definition is taken, Jiggs, an alcoholic airplane mechanic, grows disgusted with himself and flees from his host's house without even taking a bottle "since he did not want the drink right now and so when he did begin to want it, he would be at least fifteen miles away from that particular jug" [10] (Faulkner, 1935/1967, p. 119). Pouring bottles down the sink, throwing away the cigarettes or hiding the television cable are often triggered by such events. Unfortunately, these tactics are rarely successful: there are non-stop stores everywhere.

We also make deals with ourselves. If we finish the paper by June 15th, for example, we can go hiking but otherwise not. For some people these bargains work, for others they do not. For instance, one of us always takes Sunday off rather than Saturday, because if he takes Saturday off there is no way to punish himself if he then takes Sunday off—because Saturday has already been wasted—but if he wastes Saturday when he is supposed to work he can always punish himself by working on Sunday. How such intrapersonal strategies (essentially only reframings) can work baffles us, but many people do similar things, such as rewarding themselves for a job well done. Schelling (1984b, 1992) describes other such deals. He observes that we often do the equivalent of blackmailing ourselves. People who want to quit smoking make a big show of their efforts, so that if they fail they will have humiliated themselves in front of all their friends. The deadline, already mentioned, is an example of self-blackmail. If we miss the deadline (or, at least, miss it by a lot—most deadlines turn out to be flexible even though we would be better off if they weren't), then we lose credibility in front of those people we most want to impress.

The means that people use to manipulate or constrain their future behavior are virtually identical to the means that they use to manipulate or constrain the behavior of others. If society wants people to do X rather than Y, and it suspects that they will do Y if left to their own devices, society either increases the attractiveness of X by rewarding its choice, decreases the attractiveness of Y by charging for (or punishing) its choice, or else makes Y unavailable. These are the only controls that work for social choice, and they are the only controls that work for personal choice.

Notes

[1] Dawes (1980) sums up the essential characteristics of social dilemmas in the following way: "(a) each individual receives a higher payoff for a socially defecting choice (e.g. having additional children, using all the energy available, polluting his or her neighbors) than for a socially coopertive choice, no matter what the other individuals in society do, but (b) all individuals are better off if all cooperate than if all defect" (p. 169). Social dilemmas can also be conceptualized as n-person prisoner's dilemmas. Further explication can be found in Cross and Guyer (1980), Platt (1973), and Glance and Huberman (1994).

[2] Wertenbroch (in press), Read *et al.* (in press a), and Read and Van Leeuwen (1998) have all used the terms virtue and vice in this way.

[3] Invariably, there is a problem in determining whether hyperbolic discounting causes something that can be modeled in terms of multiple selves, or if multiple selves give rise to something that can be modeled in terms of hyperbolic discounting. We cannot decide one way or the other, and

it may not be a fruitful empirical question. The reader can feel free to reinterpret any contentious statements in the preferred manner.

[4] This comparison is more apt than it may at first seem. Our future selves can be thought of as our direct descendents in much the same way as we think of our children. From an evolutionary perspective, the weight we give to future selves relative to present selves should be in proportion to the contribution that future self would make to our reproductive fitness.

[5] As is clear from this example, sometimes its better to be naïve. The sophisticated smoker won't even try to quit, but the naïve one will.

[6] We talk about addictive substances. There is debate about whether people can become "addicted" to such things as gambling, sex and shopping. We believe that they can. At the same time, in common with writers such as Heyman (1996) and Becker and Murphy (1988), we also believe that addiction is responsive to market forces.

[7] We can recommend Karel Reisz's film, *The Gambler*, with a semi-autobiographical screenplay by James Toback, for a vivid depiction of this addiction.

[8] Herrnstein and Prelec (1992) and Heyman (1996) both model the process of becoming addicted as a commons-like dilemma, in which individual selves choose between consuming an addictive substance (which is locally more desirable), and not consuming it (which is globally desirable). Our account is similar to theirs, and differs primarily in that we incorporate the idea that addiction, along with other taste change dilemmas, is a one-way street (see Kirby, 1996). That is why our model gets three figures (Figures 2a–c), while theirs only requires one.

[9] There is a third category of adolescent who takes up an instrument, usually the guitar, who is motivated by social factors.

[10] Readers of this superb book will know that Jiggs does not, in fact, get 15 *meters* before his resolve changes—but by then it is too late.

References

AINSLIE, G. (1975). Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463–496.

AINSLIE, G. (1992). *Picoeconomics: the strategic interaction of successive motivational states within the person.* New York: Cambridge University Press.

AKERLOF, G. (1991). Procrastination and obedience. *American Economic Review, 81*, 1–19.

BECKER, G. & MURPHY, K.M. (1988). A theory of rational addiction. *Journal of Political Economy, 96*, 675–700.

BRATMAN, M.E. (1996). Planning and temptation. In L. MAY, M. FRIEDMAN & A. CLARK (Eds) *Mind and morals: essays on cognitive science and ethics.* Cambridge, MA: MIT Press.

CROSS, J.G. & GUYER, M.J. (1980). *Social traps.* Ann Arbor, MI: University of Michigan Press.

DAWES, R. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169–193.

ELSTER, J. (1986). *The multiple self.* Cambridge: Cambridge University Press.

FAULKNER, W. (1935/1967) *Pylon.* New York: The Modern Library.

GLANCE, N.S. & HUBERMAN, B.A. (1994). The dynamics of social dilemmas. *Scientific American.*

HARDIN, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

HERRNSTEIN, R.J. (1997). *The matching law.* New York & Cambridge: Sage & Harvard University Press.

HERRNSTEIN, R.J. & PRELEC, D. (1992). A theory of addiction. In G. LOEWENSTEIN & J. ELSTER (Eds) *Choice over time.* New York: Sage.

HERRNSTEIN, R.J., LOEWENSTEIN, G.F., PRELEC, D. & VAUGHAN JR., W. (1993). Utility maximization and melioration: internalities in individual choices. *Journal of behavioral decision making, 6*, 149–185.

HEYMAN, G.M. (1996). Resolving the contradictions of addiction. *Behavioral and Brain Sciences*, 19, 561–610.

KAHNEMAN, D., WAKKER, P. & SARIN, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics, 112*, 375–406.

KIRBY, K.N. (1996). Future directions for the melioration model of addiction. *Behavioral and Brain Sciences, 19*, 583.

LAIBSON, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 112,* 443–477.

LOEWENSTEIN, G. (1996). Out of control: visceral influences on behavior. *Organizational behavior and human decision processes, 65,* 272–292.

LOEWENSTEIN, G. (in press). A visceral account of addiction. In J. ELSTER & O.J. SKOG (Eds) *Getting hooked: rationality and addiction.* Cambridge: Cambridge University Press.

LOEWENSTEIN, G., NAGIN, D & PATERNOSTER, R. (1997). The effect of sexual arousal on predictions of sexual forcefulness. *Journal of Research in Crime and Delinquency, 34,* 443–473.

MAZUR, J.E. (1987). An adjustment procedure for studying delayed reinforcement. In: M.L. COMMONS, J.E. MAZUR, J.A. NEVIN & H. RACHLIN (Eds) *Quantitative analyses of behavior V: the effect of delay and of intervening events on reinforcement value.* Hillsdale, NJ: Erlbaum.

MELE, A.R. (1996). Addiction and self-control. *Behavior and Philosophy, 24,* 99–117.

NOZICK, R. (1993). *The nature of rationality.* Princeton, NJ: Princeton University Press.

O'DONOGHUE, E. & RABIN, M. (1997). Incentives for procrastinators. Working paper. Berkeley: Department of Economics, University of California.

O'DONOGHUE, T. & RABIN, M. (in press). The economics of immediate gratification. *Journal of Behavioral Decision Making.*

PLATT, J. (1973). Social traps. *American Psychologist, 28,* 641–651.

READ, D. & VAN LEEUWEN, B. (1998). Predicting hunger: the effects of time and appetite on. *Organizational Behavior and Human Decision Processes, 76,* 189–205.

READ, D., LOEWENSTEIN, G. & KALYANARAMAN, S. (in press a). Mixing virtue and vice: the combined effect of hyperbolic discounting and diversification. *Journal of Behavioral Decision Making.*

READ, D., LOEWENSTEIN, G. & RABIN, M. (in press b). Choice bracketing. *Journal of Risk and Uncertainty.*

SCHELLING, T.C. (1978). *Micromotives and macrobehavior.* New York: Norton.

SCHELLING, T.C. (1984a). *Choice and consequence.* Cambridge, MA: Harvard University Press.

SCHELLING, T.C. (1984b). Self-command in practice, in policy, and in a theory of rational choice. *AEA Papers and Proceedings, 74,* 1–11.

SCHELLING, T.C. (1992). Self-command: a new discipline. In G. LOEWENSTEIN & J. ELSTER (Eds) *Choice over time.* New York: Sage.

SMITH, J.M. (1975/1993). *The theory of evolution.* Cambridge: Cambridge University Press.

STROTZ, R.H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies, 23,* 165–180.

THALER, R. & SHEFRIN, H.M. (1981). An economic theory of self-control. *Journal of Political Economy, 89,* 392–406.

VARIAN, H.R. (1996). *Intermediate microeconomics, a modern approach.* New York: Norton.

WERTENBROCH, K. (in press). Consumption self-control via purchase quantity rationing. *Marketing Science.*

WILSON, J.Q. (1995). *The moral sense.* New York: Free Press.